

Human-Centered Explainable AI (XAI): From Algorithms to User Experiences

Year: 2021 | Citations: 277 | Authors: Q. Liao, Microsoft Research, Canada Kush, R. Varshney, Kush R. Varshney

Abstract

In recent years, the field of explainable AI (XAI) has produced a vast collection of algorithms, providing a useful toolbox for researchers and practitioners to build XAI applications. With the rich application opportunities, explainability is believed to have moved beyond a demand by data scientists or researchers to comprehend the models they develop, to an essential requirement for people to trust and adopt AI deployed in numerous domains. However, explainability is an inherently human-centric property and the field is starting to embrace human-centered approaches. Human-computer interaction (HCI) research and user experience (UX) design in this area are becoming increasingly important. In this chapter, we begin with a high-level overview of the technical landscape of XAI algorithms, then selectively survey our own and other recent HCI works that take human-centered approaches to design, evaluate, and provide conceptual and methodological tools for XAI. We ask the question "what are human-centered approaches doing for XAI" and highlight three roles that they play in shaping XAI technologies by helping navigate, assess and expand the XAI toolbox: to drive technical choices by users' explainability needs, to uncover pitfalls of existing XAI methods and inform new methods, and to provide conceptual frameworks for human-compatible XAI.